Quantile Regression: Inference

Roger Koenker

CEMMAP and University of Illinois, Urbana-Champaign

Les Diablerets 3-6 February 2013

Inference for Quantile Regression

- Inference for the Sample Quantiles
- QR Inference in iid Error Models*
- QR Inference in Heteroscedastic Error Models*
- Classical Rank Tests and the Quantile Regression Dual*
- Inference on the Quantile Regression Process*

* Skimmed very lightly in favor of the first DIY in R session.

What determines the precision of sample quantiles?

For random samples from a continuous distribution, F, the sample quantiles, $\hat{F}_n^{-1}(\tau)$ are consistent, by the Glivenko-Cantelli theorem.
Rates of convergence and precision are governed by the density near the quantile of interest, if it exists.
Note that differentiating the identity: $F(F^{-1}(t)) = t$, yields,

$$\frac{d}{dt}F(F^{-1}(t)) = f(F^{-1}(t))\frac{d}{dt}F^{-1}(t) = 1$$

thus, provided of course that $f(F^{-1}(t)) > 0$,

$$\frac{d}{dt}F^{-1}(t) = \frac{1}{f(F^{-1}(t))}$$

So, limiting normality of $\hat{F}_n$ and the δ-method imply limiting normality of the sample quantiles with $\sqrt{n}$ rate and variance proportional to $f^{-2}(F^{-1}(t))$.

Inference for the Sample Quantiles

Minimizing $\sum_{i=1}^{n} \rho_\tau(y_i - \xi)$ consider

$$g_n(\xi) = -n^{-1} \sum_{i=1}^{n} \psi_\tau(y_i - \xi) = n^{-1} \sum_{i=1}^{n} (I(y_i < \xi) - \tau).$$

By convexity of the objective function,

$$\{\hat{\xi}_\tau > \xi\} \Leftrightarrow \{g_n(\xi) < 0\}$$

and the DeMoivre-Laplace CLT yields, expanding F,

$$\sqrt{n}(\hat{\xi}_\tau - \xi) \rightsquigarrow \mathcal{N}(0, \omega^2(\tau, F))$$

where $\omega^2(\tau, F) = \tau(1-\tau)/f^2(F^{-1}(\tau))$. Classical Bahadur-Kiefer representation theory provides further refinement of this result.

Some Gory Details

Instead of a fixed $\xi = F^{-1}(\tau)$ consider,

$$\mathbb{P}\{\hat{\xi}_n > \xi + \delta/\sqrt{n}\} = \mathbb{P}\{g_n(\xi + \delta/\sqrt{n}) < 0\}$$

where $g_n \equiv g_n(\xi + \delta/\sqrt{n})$ is a sum of iid terms with

$$\begin{aligned}
\mathbb{E}g_n &= \mathbb{E}n^{-1}\sum_{i=1}^{n}(I(y_i < \xi + \delta/\sqrt{n}) - \tau) \\
&= F(\xi + \delta/\sqrt{n}) - \tau \\
&= f(\xi)\delta/\sqrt{n} + o(n^{-1/2}) \\
&\equiv \mu_n\delta + o(n^{-1/2})
\end{aligned}$$

$$\mathbb{V}g_n = \tau(1-\tau)/n + o(n^{-1}) \equiv \sigma_n^2 + o(n^{-1}).$$

Thus, by (a triangular array form of) the DeMoivre-Laplace CLT,

$$\mathbb{P}(\sqrt{n}(\hat{\xi}_n - \xi) > \delta) = \Phi((0 - \mu_n\delta)/\sigma_n) \equiv 1 - \Phi(\omega^{-1}\delta)$$

where $\omega = \mu_n/\sigma_n = \sqrt{\tau(1-\tau)}/f(F^{-1}(\tau))$.

Finite Sample Theory for Quantile Regression

Let $h \in \mathcal{H}$ index the $\binom{n}{p}$ p-element subsets of $\{1, 2, \ldots, n\}$ and $X(h), y(h)$ denote corresponding submatrices and vectors of X and y.

Lemma: $\hat{\beta} = b(h) \equiv X(h)^{-1} y(h)$ is the τth regression quantile iff $\xi_h \in \mathcal{C}$ where

$$\xi_h = \sum_{i \notin h} \psi_\tau(y_i - x_i \hat{\beta}) x_i^\top X(h)^{-1},$$

$\mathcal{C} = [\tau - 1, \tau]^p$, and $\psi_\tau(u) = \tau - I(u < 0)$.

Theorem: (KB, 1978) In the linear model with iid errors, $\{u_i\} \sim F, f$, the density of $\hat{\beta}(\tau)$ is given by

$$g(b) = \sum_{h \in \mathcal{H}} \prod_{i \in h} f(x_i^\top (b - \beta(\tau)) + F^{-1}(\tau)) \cdot P(\xi_h(b) \in C) |\det(X(h))|$$

Asymptotic behavior of $\hat{\beta}(\tau)$ follows by (painful) consideration of the limiting form of this density, see also Knight and Goh (ET, 2009).

Asymptotic Theory of Quantile Regression I

In the classical linear model,

$$y_i = x_i \beta + u_i$$

with u_i iid from $\mathrm{d}fF$, with density $f(u) > 0$ on its support $\{u | 0 < F(u) < 1\}$, the joint distribution of $\sqrt{n}(\hat{\beta}_n(\tau_i) - \beta(\tau_i))_{i=1}^m$ is asymptotically normal with mean 0 and covariance matrix $\Omega \otimes D^{-1}$. Here $\beta(\tau) = \beta + F_u^{-1}(\tau) e_1, e_1 = (1, 0, \ldots, 0)^\top$, $x_{1i} \equiv 1, n^{-1} \sum x_i x_i^\top \to D$, a positive definite matrix, and

$$\Omega = ((\tau_i \wedge \tau_j - \tau_i \tau_j)/(f(F^{-1}(\tau_i)) f(F^{-1}(\tau_j)))_{i,j=1}^m.$$

Asymptotic Theory of Quantile Regression II

When the response is conditionally independent over i, but not identically distributed, the asymptotic covariance matrix of $\zeta(\tau) = \sqrt{n}(\hat{\beta}(\tau) - \beta(\tau))$ is somewhat more complicated. Let $\xi_i(\tau) = x_i \beta(\tau)$, $f_i(\cdot)$ denote the corresponding conditional density, and define,

$$
\begin{aligned}
J_n(\tau_1, \tau_2) &= (\tau_1 \wedge \tau_2 - \tau_1 \tau_2) n^{-1} \sum_{i=1}^{n} x_i x_i^\top, \\
H_n(\tau) &= n^{-1} \sum x_i x_i^\top f_i(\xi_i(\tau)).
\end{aligned}
$$

Under mild regularity conditions on the $\{f_i\}$'s and $\{x_i\}$'s, we have joint asymptotic normality for $(\zeta(\tau_i), \ldots, \zeta(\tau_m))$ with covariance matrix

$$
V_n = (H_n(\tau_i)^{-1} J_n(\tau_i, \tau_j) H_n(\tau_j)^{-1})_{i,j=1}^{m}.
$$

Making Sandwiches

The crucial ingredient of the QR Sandwich is the quantile density function $f_i(\xi_i(\tau))$, which can be estimated by a difference quotient. Differentiating the identity: $F(Q(t)) = t$ we get

$$s(t) = \frac{dQ(t)}{dt} = \frac{1}{f(Q(t))}$$

sometimes called the "sparsity function" so we can compute

$$\hat{f}_i(x_i^\top \hat{\beta}(\tau)) = 2h_n/(x_i^\top(\hat{\beta}(\tau + h_n) - \hat{\beta}(\tau - h_n))$$

with $h_n = O(n^{-1/3})$. Prudence suggests a modified version:

$$\tilde{f}_i(x_i^\top \hat{\beta}(\tau)) = \max\{0, \hat{f}_i(x_i^\top \hat{\beta}(\tau))\}$$

Various other strategies can be employed including a variety of bootstrapping options. More on this in the first lab session.

Rank Based Inference for Quantile Regression

- Ranks play a fundamental *dual* role in QR inference.
- Classical rank tests for the p-sample problem extended to regression
- Rank tests play the role of Rao (score) tests for QR.

Two Sample Location-Shift Model

$$X_1, \ldots, X_n \sim F(x) \qquad \text{"Controls"}$$

$$Y_1, \ldots, Y_m \sim F(x - \theta) \qquad \text{"Treatments"}$$

Hypothesis:

$$H_0: \quad \theta = 0$$

$$H_1: \quad \theta > 0$$

The Gaussian Model $F = \Phi$

$$T = (\bar{Y}_m - \bar{X}_n)/\sqrt{n^{-1} + m^{-1}}$$

UMP Tests:

$$\text{critical region} \{T > \Phi^{-1}(1 - \alpha)\}$$

Wilcoxon-Mann-Whitney Rank Test

Mann-Whitney Form:

$$S = \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_j > X_i)$$

Heuristic: If treatment responses are larger than controls for most pairs (i, j), then H_0 should be rejected.
Wilcoxon Form: Set $(R_1, \ldots, R_{n+m}) = \text{Rank}(Y_1, \ldots, Y_m, X_1, \ldots X_n)$,

$$W = \sum_{j=1}^{m} R_j$$

Proposition: $S = W - m(m+1)/2$ so Wilcoxon and Mann-Whitney tests are equivalent.

Pros and Cons of the Transformation to Ranks

Thought One:
Gain: Null Distribution is independent of F.
Loss: Cardinal information about data.

Thought Two:
Gain: Student t-test has quite accurate size provided $\sigma^2(F) < \infty$.
Loss: Student t-test uses cardinal information badly for long-tailed F.

Asymptotic Relative Efficiency of Wilcoxon versus Student t-test

Pitman (Local) Alternatives: $H_n : \theta_n = \theta_0/\sqrt{n}$
$(\text{t-test})^2 \rightsquigarrow \chi_1^2(\theta_0^2/\sigma^2(F))$
$(\text{Wilcoxon})^2 \rightsquigarrow \chi_1^2(12\theta_0^2(\int f^2)^2)$

$$\text{ARE}(W, t, F) = 12\sigma^2(F)[\int f^2(x)dx]^2$$

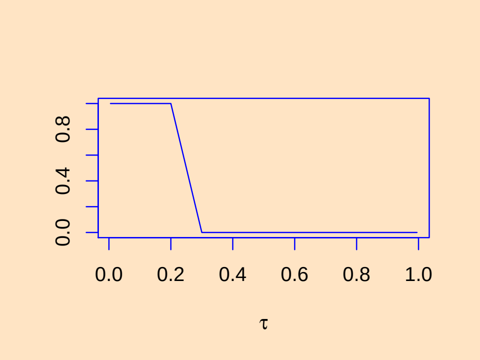
F	N	U	Logistic	DExp	LogN	t_2
ARE	.955	1.0	1.1	1.5	7.35	∞

Theorem (Hodges-Lehmann) For all F, $ARE(W, t, F) \geqslant .864$.

Hájek 's Rankscore Generating Functions

Let $Y_1, \ldots, Y_n$ be a random sample from an absolutely continuous df F with associated ranks $R_1, \ldots, R_n$, Hájek 's rank generating functions are:

$$\hat{a}_i(t) = \begin{cases} 1 & \text{if } t \leqslant (R_i - 1)/n \\ R_i - tn & \text{if } (R_i - 1)/n \leqslant t \leqslant R_i/n \\ 0 & \text{if } R_i/n \leqslant t \end{cases}$$



Linear Rank Statistics Asymptotics

Theorem (Hájek (1965)) Let $c_n = (c_{1n}, \ldots, c_{nn})$ be a triangular array of real numbers such that

$$\max_i (c_{in} - \bar{c}_n)^2 / \sum_{i=1}^{n} (c_{in} - \bar{c}_n)^2 \to 0.$$

Then

$$\begin{aligned} Z_n(t) &= \left(\sum_{i=1}^{n} (c_{in} - \bar{c}_n)^2\right)^{-1/2} \sum_{j=1}^{n} (c_{jn} - \bar{c}_n) \hat{a}_j(t) \\ &\equiv \sum_{j=1}^{n} w_j \hat{a}_j(t) \end{aligned}$$

converges weakly to a Brownian Bridge, i.e., a Gaussian process on $[0, 1]$ with mean zero and covariance function $\mathrm{Cov}(Z(s), Z(t)) = s \wedge t - st$.

Some Asymptotic Heuristics

The Hájek functions are approximately indicator functions

$$\hat{a}_i(t) \approx I(Y_i > F^{-1}(t)) = I(F(Y_i) > t)$$

Since $F(Y_i) \sim U[0,1]$, linear rank statistics may be represented as

$$\int_0^1 \hat{a}_i(t) d\varphi(t) \approx \int_0^1 I(F(Y_i) > t) d\varphi(t) = \varphi(F(Y_i)) - \varphi(0)$$

$$\begin{aligned} \int_0^1 Z_n(t) d\varphi(t) &= \sum w_i \int \hat{a}_i(t) d\varphi(t) \\ &= \sum w_i \varphi(F(Y_i)) + o_p(1), \end{aligned}$$

which is asymptotically distribution free, i.e. independent of F.

Duality of Ranks and Quantiles

Quantiles may be *defined* as

$$\hat{\xi}(\tau) = \text{argmin} \sum \rho_\tau(y_i - \xi)$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$. This can be formulated as a linear program whose dual solution

$$\hat{a}(\tau) = \text{argmax}\{y^\top a | 1_n^\top a = (1-\tau)n, a \in [0,1]^n\}$$

generates the Hájek rankscore functions.

Reference: Gutenbrunner and Jurečková (1992).

Regression Quantiles and Rank Scores:

$$\hat{\beta}_n(\tau) = \text{argmin}_{b \in R^p} \sum \rho_\tau(y_i - x_i^\top b)$$

$$\hat{a}_n(\tau) = \text{argmax}_{a \in [0,1]^n}\{y^\top a | X^\top a = (1-\tau)X^\top 1_n\}$$

$x^\top \hat{\beta}_n(\tau)$ — Estimates $Q_Y(\tau|x)$
Piecewise constant on $[0,1]$.
For $X = 1_n,\ \hat{\beta}_n(\tau) = \hat{F}_n^{-1}(\tau)$.

$\{\hat{a}_i(\tau)\}_{i=1}^n$ — Regression rankscore functions
Piecewise linear on $[0,1]$.
For $X = 1_n,\ \hat{a}_i(\tau)$ are Hajek rank generating functions.

Regression Rankscore "Residuals"

The Wilcoxon rankscores,

$$\tilde{u}_i = \int_0^1 \hat{a}_i(t)dt$$

play the role of quantile regression residuals. For each observation y_i they answer the question: on which quantile does y_i lie? The $\tilde{u}_i$ satisfy an orthogonality restriction:

$$X^\top \tilde{u} = X^\top \int_0^1 \hat{a}(t)dt = n\bar{x}\int_0^1 (1-t)dt = n\bar{x}/2.$$

This is something like the $X^\top \hat{u} = 0$ condition for OLS. Note that if the X is "centered" then $\bar{x} = (1, 0, \cdots, 0)$. The $\tilde{u}$ vector is approximately uniformly "distributed;" in the one-sample setting $u_i = (R_i + 1/2)/n$ so they are obviously "too uniform."

Regression Rank Tests

$$Y = X\beta + Z\gamma + u$$

$$H_0 : \gamma = 0 \text{ versus } H_n : \gamma = \gamma_0/\sqrt{n}$$

Given the regression rank score process for the restricted model,

$$\hat{a}_n(\tau) = \text{argmax}\left\{ Y^\top a \,|\, X^\top a = (1-\tau)X^\top 1_n \right\}$$

A test of H_0 is based on the linear rank statistics,

$$\hat{b}_n = \int_0^1 \hat{a}_n(t)\, d\varphi(t)$$

Choice of the score function φ permits test of location, scale or (potentially) other effects.

Regression Rankscore Tests

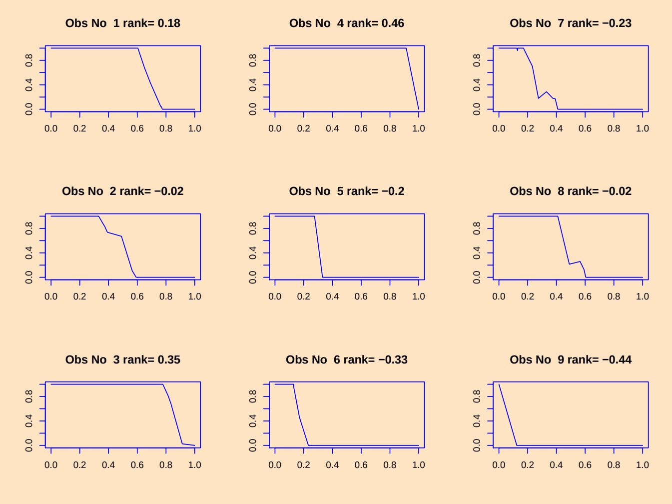
Theorem: (Gutenbrunner, Jurečková , Koenker and Portnoy) Under H_n and regularity conditions, the test statistic $T_n = S_n^\top Q_n^{-1} S_n$ where $S_n = (Z - \hat{Z})^\top \hat{b}_n, \quad \hat{Z} = X(X^\top X)^{-1} X^\top Z$, $Q_n = n^{-1}(Z - \hat{Z})^\top Z - \hat{Z})$

$$T_n \rightsquigarrow \chi_q^2(\eta)$$

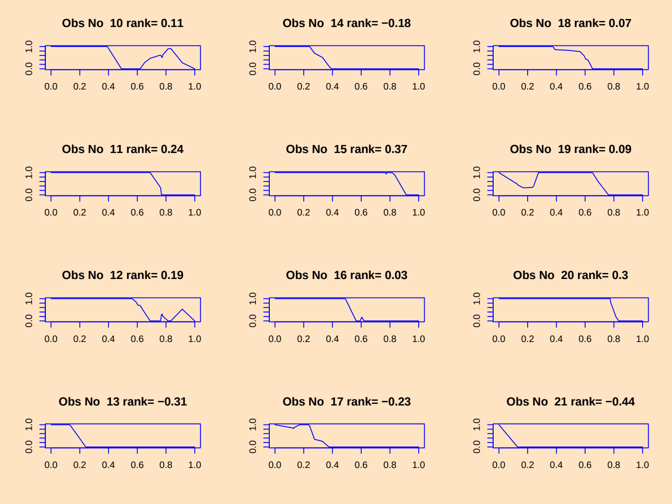
where

$$\begin{aligned} \eta^2 &= \omega^2(\varphi, F)\gamma_0^\top Q \gamma_0 \\ \omega(\varphi, F) &= \int_0^1 f(F^{-1}(t))\, d\varphi(t) \end{aligned}$$

Regression Rankscores for Stackloss Data



Regression Rankscores for Stackloss Data



Inversion of Rank Tests for Confidence Intervals

For the scalar γ case and using the score function

$$\varphi_\tau(t) = \tau - I(t < \tau)$$

$$\hat{b}_{ni} = -\int_0^1 \varphi_\tau(t) d\hat{a}_{ni}(t) = \hat{a}_{ni}(\tau) - (1-\tau)$$

where $\bar{\varphi} = \int_0^1 \varphi_\tau(t)dt = 0$ and $A^2(\varphi_\tau) = \int_0^1 (\varphi_\tau(t) - \bar{\varphi})^2 dt = \tau(1-\tau)$. Thus, a test of the hypothesis $H_0 : \gamma = \xi$ may be based on $\hat{a}_n$ from solving,

$$\max\{(y - x_2\xi)^\top a | X_1^\top a = (1-\tau)X_1^\top 1, a \in [0,1]^n\} \tag{1}$$

and the fact that

$$S_n(\xi) = n^{-1/2} x_2^\top \hat{b}_n(\xi) \rightsquigarrow \mathcal{N}(0, A^2(\varphi_\tau) q_n^2) \tag{2}$$

Inversion of Rank Tests for Confidence Intervals

That is, we may compute

$$T_n(\xi) = S_n(\xi)/(A(\varphi_\tau)q_n)$$

where $q_n^2 = n^{-1}x_2^\top(I - X_1(X_1^\top X_1)^{-1}X_1^\top)x_2$. and reject H_0 if $|T_n(\xi)| > \Phi^{-1}(1 - \alpha/2)$.

Inverting this test, that is finding the interval of ξ's such that the test fails to reject. This is a quite straightforward parametric linear programming problem and provides a simple and effective way to do inference on individual quantile regression coefficients. Unlike the Wald type inference it delivers asymmetric intervals. This is the default approach to parametric inference in `quantreg` for problems of modest sample size.

Inference on the Quantile Regression Process

Using the quantile score function, $\varphi_\tau(t) = \tau - I(t < \tau)$ we can consider the quantile rankscore process,

$$T_n(\tau) = S_n(\tau)^\top Q_n^{-1} S_n(\tau)/(\tau(1-\tau)).$$

where

$$\begin{aligned}
S_n &= n^{-1/2}(X_2 - \hat{X}_2)^\top \hat{b}_n, \\
\hat{X}_2 &= X_1(X_1^\top X_1)^{-1} X_1^\top X_2, \\
Q_n &= (X_2 - \hat{X}_2)^\top (X_2 - \hat{X}_2)/n, \\
\hat{b}_n &= (-\int \varphi(t) d\hat{a}_{in}(t))_{i=1}^n,
\end{aligned}$$

Inference on the Quantile Regression Process

Theorem: (K & Machado) Under $H_n : \gamma(\tau) = O(1/\sqrt{n})$ for $\tau \in (0, 1)$ the process $T_n(\tau)$ converges to a non-central Bessel process of order $q = \dim(\gamma)$. Pointwise, T_n is non-central χ^2.

Related Wald and LR statistics can be viewed as providing a general apparatus for testing goodness of fit for quantile regression models. This approach is closely related to classical p-dimensional goodness of fit tests introduced by Kiefer (1959).

When the null hypotheses under consideration involve unknown nuisance parameters things become more interesting. In Koenker and Xiao (2001) we consider this "Durbin problem" and show that the elegant approach of Khmaladze (1981) yields practical methods.

Four Concluding Comments about Inference

- Asymptotic inference for quantile regression poses some statistical challenges since it involves elements of nonparametric density estimation, but this shouldn't be viewed as a major obstacle.
- Classical rank statistics and Hájek 's rankscore process are closely linked via Gutenbrunner and Jurečková 's regression rankscore process, providing an attractive approach to many inference problems while avoiding density estimation.
- Inference on the quantile regression process can be conducted with the aid of Khmaladze's extension of the Doob-Meyer construction.
- Resampling offers many further lines of development for inference in the quantile regression setting.